On Delta-Method of Moments and Probabilistic Sums*

Jacek Cichoń¶ Zbigniew Gołębiewski¶ Marcin Kardas¶ Marek Klonowski¶

Abstract
We discuss a general framework for determining asymptotics of the expected value of random variables of the form $f(X)$ in terms of a function f and central moments of the random variable X. This method may be used for approximation of entropy, inverse moments, and some statistics of discrete random variables useful in analysis of some randomized algorithms. Our approach is based on some variant of the Delta Method of Moments. We formulate a general result for an arbitrary distribution and next we show its specific extension to random variables which are sums of identically distributed independent random variables. Our method simplifies previous proofs of results of several authors and can be automated to a large extend. We apply our method to the binomial, negative binomial, Poisson and hypergeometric distribution. We extend the class of function for which our method is applicable for some subclass of exponential functions and double exponential function for some cases.
Keywords: Delta Method of Moment, Binomial Sums, entropy, binomial distribution, Poisson distribution, negative-binomial distribution, hypergeometric distribution

1 Introduction

Let X be a random variable with expected value μ and let $f : \mathbb{R} \to \mathbb{R}$ be function which is differentiable in some neighborhood of μ. Our goal is to investigate the expected value of the random variable $Y = f(X)$, i.e., we want to control the difference between the expected value of Y and the number $f(\mu)$.

In this paper we extend results from papers [1], [2] and [3]. A previous approach of two authors of this paper (see [4]) to the same problems was based on Bernstein polynomials. The approach from this paper, based on the Delta Method of Moments, is more general and can be automated more easily than the previous one. Our method works, roughly speaking, for function that has an analytic extensions to a function of a polynomial growth in a right half-plane of the complex plane. Due to some algebraic properties of analyzed distributions we extend the class of applicable function to exponential and even doubly exponential functions in some cases (Lemma 3.1, Lemma 3.2).

Our main results presented in this paper are based on the Delta Method. In many cases this tool turned out to be very useful for finding approximation of expected value of functions of random variables in when direct calculations are hard. The first fairly rigorous formulation of this method was done by H. Cramer in 1946 [5], who used it for estimation of moments of functions of samples from populations (see [6]).

Let us note that in [7] during an analysis of some distributed algorithm for Ad Hoc networks a formula $\sum_{k\geq 1}\log(k)\binom{n}{k}p^k(1-p)^k = \log n + (p-1)/(2np) + \mathrm{O}\left(1/n^2\right)$ was used in order to build an estimator for the network size. Using methods presented in this paper this kind of formulas can be derived without difficulty. This is one of typical application of methods from this paper for construction and analysis of algorithms.

1.1 Paper Organization In Section 2 we provide tools that can be used for estimating the expectation of $f(X)$ (Theorem 2.1 and Theorem 2.2). In next sections we discuss its applications for the binomial, Poisson, negative-binomial and hypergeometric distributions.

1.2 Notations The expected value of a random variable X is denoted by $\mathrm{E}[X]$ and its variance by $\mathrm{var}[X]$. The kth central moment of a random variable X is denoted by $\mu_k(X)$, i.e. $\mu_k(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$. We also put $|\mu_k|(X) = \mathrm{E}\left[|X - \mathrm{E}[X]|^k\right]$.

We use the abbreviation i.i.d. for "independent and identically distributed" collection or sequence of random variables.

By $\mathrm{Ber}(p)$ and $\mathrm{Geo}(p)$ we denote the Bernoulli and the Geometric distributions, respectively with parameter p. By $\mathrm{Bin}(n,p)$ we denote the binomial distribution with parameters n and p. Similarly, by $\mathrm{NB}(k,p)$ we denote the negative binomial distribution with parameters k and p.

We denote by $\mathrm{HGeo}(n, n_1, n_2)$ the hypergeometric distribution with parameters n, n_1 and n_2. I.e., if X has $\mathrm{HGeo}(n, n_1, n_2)$ distribution, then $\Pr(X = k) = \binom{n_2}{k}\binom{n-n_2}{n_1-k}\Big/\binom{n}{n_1}$. In applications parameter n denotes the size of population, n_1 denotes the size of a sample, n_2 is the number of marked elements in the population and X denotes the number of marked elements in the sample.

*Supported by grant number 350600/NN206369739 of the Polish Ministry of Science and Higher Education
¶Institute of Mathematics and Computer Science, Wroclaw University of Technology

We denote by $f^{(k)}(x)$ the kth derivative of the function f at point x. In the case $k=0$ we put $f^{(0)}=f$. For a real function f and $A \subseteq \mathbb{R}$ we put

$$\mathrm{M}_A^{(t)}(f) = \sup\{|f^{(t)}(x)| : x \in A\} \ .$$

Observe that $\mathrm{M}_A^{(0)}(f) = \sup\{|f(x)| : x \in A\}$.

2 Delta Method of Moments

The Delta Method is a natural technique for approximating the moments of functions of random variables based on the Taylor formula. We prove in this section some variant of the Delta Method which we found useful for our further considerations. Another, less specialized, variant of this method can be found in [8] Sec. 5.3.1.

THEOREM 2.1. *Let X be a random variable and let $s \geq 1$ be a natural number. Let $\mu = \mathrm{E}[X]$. Suppose that $\mu_{2(s-1)}(X) < \infty$. Let $-\infty \leq m \leq a \leq \mu \leq b \leq M \leq \infty$ be such that $\Pr(m \leq X \leq M) = 1$. Suppose that f is s-times differentiable on (a,b). Then*

$$\mathrm{E}[f(X)] = f(\mu) + \sum_{k=2}^{s-1} \frac{f^{(k)}(\mu)}{k!}\mu_k(X) + R_{s;a,b},$$

where $|R_{s;a,b}| \leq U + W + V$ and

1. $U = \frac{|\mu_s|(X)}{s!}\mathrm{M}_{(a,b)}^{(s)}(f)$,
2. $W = \sqrt{\Pr(X \notin (a,b))}\sum_{k=0}^{s-1}\left(\left|\frac{f^{(k)}(\mu)}{k!}\right|\sqrt{\mu_{2k}(X)}\right)$,
3. $V = \Pr(X \notin (a,b)) \cdot \mathrm{M}_{[m,M]\setminus(a,b)}^{(0)}(f)$.

Proof. Let A denote the event that $a < X < b$ and let $B = \Omega \setminus A$. Then

$$\mathrm{E}[f(X)] = \int_A f(X)dP + \int_B f(X)dP \ , \tag{2.1}$$

and

$$\left|\int_B f(X)dP\right| \leq \Pr(B) \cdot \mathrm{M}_{[m,M]\setminus(a,b)}^{(0)}(f) \ . \tag{2.2}$$

From the Taylor formula at the point μ with the Lagrange type of the remainder we get

$$f(x) = \sum_{k=0}^{s-1} \frac{f^{(k)}(\mu)}{k!}(x-\mu)^k + \frac{f^{(s)}(x^*)}{s!}(x-\mu)^s \ ,$$

where $x^* = \mu + \theta(x-\mu)$ for some $\theta \in (0,1)$. Notice that if $x \in (a,b)$ then $x^* \in (a,b)$, too. Hence

$$\int_A f(X)dP = \sum_{k=0}^{s-1} \frac{f^{(k)}(\mu)}{k!}\int_A (X-\mu)^k \, dP + T_{s;a,b} \ , \tag{2.3}$$

where

$$|T_{s;a,b}| \leq \frac{1}{s!}\int_A \left|f^{(s)}(X^*)\right| |X-\mu|^s \, dP \leq$$
$$\frac{1}{s!}\mathrm{M}_{(a,b)}^{(s)}(f)\int_A |X-\mu|^s \, dP \leq$$
$$\frac{1}{s!}\mathrm{M}_{(a,b)}^{(s)}(f)\int_\Omega |X-\mu|^s \, dP = \frac{1}{s!}\mathrm{M}_{(a,b)}^{(s)}(f) \cdot |\mu_s|(X) \ .$$

From the Cauchy inequality we get

$$\left|\int_B (X-\mu)^k dP\right| \leq \int_\Omega \mathbf{1}_B |X-\mu|^k dP \leq \sqrt{\int_\Omega |X-\mu|^{2k} dP}\sqrt{\Pr(B)} = \sqrt{|\mu_{2k}|(X)}\sqrt{\Pr(B)}. \tag{2.4}$$

Notice finally that $\int_A (X-\mu)^k dP = \mu_k(X) - \int_B (X-\mu)^k dP$. Therefore

$$\sum_{k=0}^{s-1} \frac{f^{(k)}(\mu)}{k!}\int_A (X-\mu)^k \, dP = \sum_{k=0}^{s-1} \frac{f^{(k)}(\mu)}{k!}\mu_k(X) - \sum_{k=0}^{s-1} \frac{f^{(k)}(\mu)}{k!}\int_B (X-\mu)^k \, dP \ . \tag{2.5}$$

Putting equalities (2.1), (2.5) and inequalities (2.2), (2.3), (2.4) together we obtain the thesis.

Notice that for $s=3$ we get from Theorem 2.1

$$\mathrm{E}[f(X)] = f(\mu) + \frac{1}{2}\mathrm{var}[X] f^{(2)}(\mu) + R_{3;a,b} \ ,$$

therefore in order to apply this theorem we need the knowledge of $\mu_1(X)$, $\mu_2(X)$, $|\mu_3|(X)$, $\mu_4(X)$, $\Pr[X \in (a,b)]$ and we need to estimate $|f^{(3)}|$ on (a,b) and $|f|$ on the set $[m,M] \setminus (a,b)$.

Before we formulate next result we prove one auxiliary lemma, which gives a bound on even moments of sums of independent random variables.

LEMMA 2.1. *Let $s \geq 1$. There exists a constant C_s such that for every sequence $X_1,\ldots,X_n$ of i.i.d. random variables such that $\mathrm{E}[X_i]=0$ we have*

$$\mathrm{E}\left[\left(\sum_{k=1}^n X_k\right)^{2s}\right] \leq C_s \cdot n^s \cdot \mu_{2s}(X_1) \ .$$

Proof. From Marcinkiewicz-Zygmund inequality (see [9]) we get

$$\mathrm{E}\left[\left(\sum_{k=1}^n X_k\right)^{2s}\right] \leq C_s \mathrm{E}\left[\left(\sum_{k=1}^n (X_k)^2\right)^s\right] \ ,$$

where C_s depends only on s. Next we use Hölder inequality with parameters $(s-1)/s$ and $1/s$ and get

$$\sum_{k=1}^{n}(X_k)^2 \le \left(\sum_{k=1}^{n}1\right)^{\frac{s-1}{s}}\left(\sum_{k=1}^{n}(X_k)^{2s}\right)^{\frac{1}{s}} = n^{\frac{s-1}{s}}\left(\sum_{k=1}^{n}(X_k)^{2s}\right)^{\frac{1}{s}},$$

so $\left(\sum_{k=1}^{n}(X_k)^2\right)^s \le n^{s-1}\sum_{k=1}^{n}(X_k)^{2s}$, and hence

$$\mathrm{E}\left[\left(\sum_{k=1}^{n}X_k\right)^{2s}\right] \le C_s n^{s-1}\mathrm{E}\left[\sum_{k=1}^{n}(X_k)^{2s}\right] = C_s n^s \mathrm{E}\left[(X_1)^{2s}\right] .$$

Now, we formulate a version of Theorem 2.1 for a random variable that is a sum of i.i.d. random variables.

THEOREM 2.2. *Let (X_k) be a sequence of i.i.d. random variables. Let $\mu = \mathrm{E}[X_1]$ and $S_n = X_1 + \cdots + X_n$ and $s > 1$. Suppose that $\mu_{4s}(X_1) < \infty$. Let $-\infty \le m \le a \le \mu \cdot n \le b \le M \le \infty$ be such that $\Pr(m \le S_n \le M) = 1$. Suppose that f is $2s$-times differentiable on (a,b). Then*

$$\mathrm{E}[f(S_n)] = f(\mu n) + \sum_{k=2}^{2s-1}\frac{f^{(k)}(\mu n)}{k!}\mu_k(S_n) + R_{2s;a,b},$$

where $|R_{2s;a,b}| \le U + W + V$ and

1. $U = n^s \cdot A_s \cdot \mathrm{M}^{(2s)}_{(a,b)}(f)$,
2. $W = B_{s,f}\sqrt{\Pr(S_n \notin (a,b))}$,
3. $V = \mathrm{M}^{(0)}_{[m,M]\setminus(a,b)}(f)\Pr(S_n \notin (a,b))$

and A_s depends only on s. The term $B_{s,f}$ may be bounded by

$$\sum_{k=0}^{2s-1} D_k n^{\frac{k}{2}}\left|f^{(k)}(\mu n)\right| ,$$

where constants $(D_k)_{k=0\ldots,2s-1}$ depends only on $2k$-th moment of X_1.

Proof. It is clear that we may assume that $\mu = 0$. We shall apply Theorem 2.1 to the random variable $S_n = X_1 + \cdots + X_n$. Notice that $\mathrm{E}[S_n] = n\mu = 0$. Therefore

$$\mathrm{E}[S_n] = \sum_{k=0}^{2s-1}\frac{f^{(k)}(\mu n)}{k!}\mu_k(S_n) + R_{2s;a,b}$$

where $|R_{2s;a,b}| \le U + W + V$ and

1. $U = \frac{\mu_{2s}(|S_n|)}{(2s)!}\mathrm{M}^{(2s)}_{(a,b)}(f)$,
2. $W = \sqrt{\Pr(S_n \notin (a,b))}\sum_{k=0}^{2s-1}\frac{f^{(k)}(\mu n)}{k!}\sqrt{\mu_{2k}(|S_n|)}$,
3. $V = \Pr(S_n \notin (a,b)) \cdot \mathrm{M}^{(0)}_{[m,M]\setminus(a,b)}(f)$.

From Lemma 2.1 we deduce that there exists a constant C_s, depending only on s such that $\mu_{2s}(|S_n|) = \mathrm{E}\left[(S_n)^{2s}\right] \le C_s n^s \mu_{2s}(X_1)$. Therefore

$$U \le \frac{C_s}{(2s)!}n^s\mu_{2s}(X_1)\mathrm{M}^{(2s)}_{(a,b)}(f) .$$

So, if we put $A_s = (C_s\mu_{2s}(X_1))/(2s)!$, then we get $U \le n^s A_s \mathrm{M}^{(2s)}_{(a,b)}(f)$.

Using once again Lemma 2.1 we deduce that there exist constants $(C_k)_{k=0,\ldots 4s-2}$, depending only on k such that $\mu_{2k}(|S_n|) \le C_k n^k \mu_{2k}(X_1)$. Therefore

$$\sum_{k=0}^{2s-1}\left|\frac{f^{(k)}(\mu n)}{k!}\right|\sqrt{\mu_{2k}(|S_n|)} \le \sum_{k=0}^{2s-1}\left|\frac{f^{(k)}(\mu n)}{k!}\right| n^{\frac{k}{2}}\sqrt{C_k}\sqrt{\mu_{2k}(X_1)} .$$

So, if we put $D_k = \sqrt{C_k}/k!\sqrt{\mu_{2k}(X_1)}$, then we get

$$W \le \sqrt{\Pr(S_n \notin (a,b))}\sum_{k=0}^{2s-1} D_k n^{\frac{k}{2}}\left|f^{(k)}(\mu n)\right| .$$

The last term V is in the required form. So the proof is done.

2.1 Example We will show one simple example of application of Theorem 2.2 for the binomial distributions. Namely, let us consider the function $f(t) = \log(t)$. We put $f(t) = 0$ if $t \le 1$. Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with Bernoulli distribution $\mathrm{Ber}(p)$, where $0 < p < 1$. Then S_n follows the Binomial distribution with parameters n and p. Notice that

$$\mathrm{E}[f(S_n)] = \sum_{k=1}^{n}\log(k)\binom{n}{k}p^k(1-p)^{n-k} .$$

We put $m = 0$, $M = n$, $a = np/2$ and $b = n$. Then $m \le S_n \le M$ and $m < a < \mathrm{E}[S_n] = np < b = M$. We apply Theorem 2.2 for $s = 2$ and we get the approximation

$$\mathrm{E}[f(S_n)] = \log(np) - \frac{1-p}{2np} + \frac{(1-p)(1-2p)}{3n^2p^2} + R_{4;a,b}.$$

We shall estimate the error term. Notice that $\Pr(S_n \notin (a,b)) = \Pr(S_n \le np/2) = \mathrm{O}\left(n^{-4}\right)^*$, hence $\mathrm{M}^{(0)}_{[m,M]}(f)\Pr(S_n \notin (a,b)) = \mathrm{O}\left(\log(n)n^{-4}\right) = \mathrm{O}\left(n^{-3}\right)$

*In fact, Chernoff bounds gives a much stronger result and $n^2 \Pr(S_n \notin (a,b)) = \mathrm{O}\left(n^{-2}\right)$. Notice that $f^{(4)}(x) = (-6)x^{-4}$ on the interval (a,b), so

$$n^2 \mathrm{M}^{(4)}_{(a,b)}(f) = n^2 \frac{6}{(np/2)^4} = \mathrm{O}\left(n^{-2}\right) \ ,$$

hence $R_{4;a,b} = \mathrm{O}\left(n^{-2}\right)$. Therefore

$$\sum_{k=1}^{n} \log(k) \binom{n}{k} p^k (1-p)^{n-k} = \log(np) - \frac{1-p}{2np} + \mathrm{O}\left(\frac{1}{n^2}\right) .$$

We get this way the first order approximation of the fourth formula from Proposition 1 of [2].

2.2 Remarks Suppose that we consider a function f which has an analytic continuation on a half-plane $\Gamma_a = \{z \in \mathbb{C} : \Re(z) \geq a\}$ for some fixed a. Assume that the function f is of polynomial growth, i.e., that $|f(z)| = \mathrm{O}\left(|z|^\alpha\right)$ when $|z| \to \infty$ in region Γ_a, where α is some fixed real number. Let (X_k) be a sequence of i.i.d. random variables such that $\mathrm{E}\left[X_1\right] = \mu > 0$ and let $S_n = X_1 + \ldots + X_n$. Then $\mathrm{E}\left[S_n\right] = n\mu$. We deduce that if $x \approx n\mu$, then $f^{(k)}(x) = \mathrm{O}\left(n^{\alpha-k}\right)$. This observation allows us to use Theorem 2.2 for a large class of functions of polynomial growth. We will illustrate this method in next sections.

3 Binomial Distribution

Let $f : \mathbb{R} \to \mathbb{R}$ be real valued function and let $p \in (0,1)$. Let Z_n be random variable with $\mathrm{Bin}(n,p)$. Then Z_n may be represented as a sum $Z_n = X_1 + \ldots + X_n$ of sequence of independent Bernoulli trials with success probability p. Let $\mathcal{B}(f; n, p) = \mathrm{E}\left[f(Z_n)\right]$.

Let $\epsilon > 0$ be such that $(1-\epsilon)p > 0$ and $(1+\epsilon)p < 1$. If we apply Theorem 2.2 for $a = (1-\epsilon)np$, $b = (1+\epsilon)np$ and $s = 3$, then we get the following slightly complicated formula:

$$\begin{gathered}
\mathrm{E}\left[Z_n\right] = \sum_k f(k) \binom{n}{k} p^k (1-p)^{n-k} = f(np) + \\
\frac{1}{24} n^2 (p-1)^2 p^2 \left(3 f^{(4)}(np) + 2(1-2p) f^{(5)}(np)\right) - \\
\frac{1}{120} n(p-1)p\Big(-24p^3 f^{(5)}(np) + 30p^2 f^{(4)}(np) + \\
36p^2 f^{(5)}(np) - 30p f^{(4)}(np) - 14p f^{(5)}(np) + \\
(20-40p) f^{(3)}(np) + 5 f^{(4)}(np) + \\
f^{(5)}(np) + 60 f''(np)\Big) + \\
R_{6;a,b}
\end{gathered}$$

The error term $R_{6;a,b}$ is divided into three parts. The part $\mathrm{M}^{(0)}_{[0,n]}(f) \Pr(S_n \notin (a,b))$ is exponentially small if the function f is of a polynomial growth. The term $n^3 \sqrt{\Pr(S_n \notin (a,b))}$ is also exponentially small. Hence, if we are able to show that $\mathrm{M}^{(6)}_{(a,b)}(f) = \mathrm{O}\left(n^\beta\right)$, then for the last term we get the estimation of $\mathrm{E}\left[Z_n\right]$ with precision $\mathrm{O}\left(n^{\beta+3}\right)$.

The above formula for $\mathrm{E}\left[Z_n\right]$ without the error term can be easily manipulated by symbolic computation packages. Using this method we can automatically derive **all examples** from [2]. In a similar way we can derive a formula from Corollary 2 from [3] for negative moments of the binomial distribution. In a completely automated way (see Sec. 3.1) we may also derive the formula for entropy of the binomial distribution $\mathrm{Bin}(n,p)$:

(3.6)

$$H = \log\sqrt{2\pi nepq} + \frac{1}{12n}\left(4 - \left(\frac{1}{p} + \frac{1}{q}\right)\right) + \mathrm{O}\left(\frac{1}{n^2}\right) ,$$

where $q = 1-p$.

The second term of expression (3.6) is in a closed form. The same term from [1] (Theorem 2) for the entropy of the binomial distribution contains a complicated infinite sum containing coefficients of the expansion of the function $\exp(x\ln(1+\lambda(\mathrm{e}^y - 1)) - \lambda xy)$ as an infinite double series of variables x and y. In [2] the entropy of the binomial distribution is calculated with accuracy of order $\mathrm{O}(n^{-1})$.

3.1 Mathematica Code The following listing shows a session with Mathematica package during which the entropy for the binomial distribution was calculated.

```
Central moment generating function
for the binomial distribution
CMGFBin[t_,n_,p_]:= (1-p + p E^t)^n/Exp[n p t]

Central ath moments of the binomial distribution
CMBin[n_,p_,a_]:= Simplify[D[CMGFBin[t,n,p],
   {t,a}]/.t->0]

Approximation Formula
AppBin[f_,n_,p_,s_]:=
  Sum[Simplify[Derivative[a][f][n p],n>1]
  CMBin[n,p,a]/a!,{a,0,2s -1}];

Entropy of the binomial distribution
H=AppBin[
 Function[x,-Log[Gamma[n+1]/(Gamma[x+1]Gamma[n-x+1])
 p^x (1-p)^(n-x)]],
 n,p,3];

Simplification
Simplify[Series[H,{n,\[Infinity],2},
  Assumptions->0<p<1],0<p<1&&n>1]

1/2 (1+Log[-2 n (-1+p) p \[Pi]])+
(1-2 p)^2/(12 (-1+p) p n)+
(-13+64 p-126 p^2+124 p^3-62 p^4)/
(24 (-1+p)^2 p^2 n^2)+O[1/n]^(5/2)

Making second term more human readable
Apart[(1-2 p)^2/((-1+p) p)]/.(-1+p)->-q
4-1/p-1/q
```

This code may be easily reused for other functions — for this purpose, it is sufficient to change the line containing the formula for entropy of the binomial distribution. The calculation of central moments of the binomial distribution can be speeded up if we use the following recursive formula

$\mu_{n,p;a+1} = p(1-p)\left(\frac{d\mu_{n,p;a}}{dp} + na\mu_{n,p;a-1}\right)$ for central moments.

It is also clear that the above code can be easily converted to work with other distributions — it is sufficient to replace the first line by the proper formula for generating function of the central moments of considered distribution.

3.2 Extensions The method described above may be extended to function of the form $f(z) = w(z) \cdot a^z$, where $a > 0$ and w is a real function for which the Delta Method from Theorem 2.2 works.

LEMMA 3.1. *Let $a > 0$, $f(z) = w(z) \cdot a^z$, $x \in (0,1)$ and $x_a = \frac{ax}{ax+1-x}$. Then $x_a \in (0,1)$ and*

$$\mathcal{B}(f; n, x) = (1 + x(a-1))^n \mathcal{B}(w; n, x_a) \ .$$

Proof. Let $a > 0$ and $x \in (0,1)$. The condition $x_a \in (0,1)$ is easy to be checked. We have

$$\frac{1}{(ax+1-x)^n}\mathcal{B}(f; n, x) = \\ \frac{1}{(ax+1-x)^n}\sum_{k=0}^{n} w(k)a^k\binom{n}{k}x^k(1-x)^{n-k} = \\ \sum_{k=0}^{n} w(k)\binom{n}{k}\left(\frac{ax}{ax+1-x}\right)^k\left(\frac{1-x}{ax+1-x}\right)^{n-k} = \\ \sum_{k=0}^{n} w(k)\binom{n}{k}(x_a)^k(1-x_a)^{n-k} = \mathcal{B}(w; n, x_a),$$

so the lemma is proved.

As an example of application of Lemma 3.1 we consider the formula from Section 2.1) for $\mathcal{B}(\ln; n, x)$ and from Lemma 3.1 get

$$\sum_{k\geq 1}\log(k)2^k\binom{n}{k}x^k(1-x)^{n-k} = \\ (1+x)^n\left(\log\left(\frac{2nx}{1+x}\right) - \frac{1-x}{4xn} + \mathrm{O}\left(\frac{1}{n^2}\right)\right) \ .$$

By a double exponential function we mean a function of the form $f(z) = a^{(b^z)}$. Let us recall that for each $a, b > 1$ and $c > 0$ we have $c^n = o\left(a^{(b^n)}\right)$, when n tends to infinity.

LEMMA 3.2. *Let $a, b > 1$, $x \in (0,1)$, $f(z) = w(z) \cdot a^{(b^z)}$, where $w(z)$ function such that $0 < c < |w(n)| \leq C^n$ for sufficiently large n. Then*

$$\mathcal{B}(f; n, x) = a^{(b^n)}x^n w(n)\left(1 + \mathrm{O}\left(\frac{D^n}{d^{(b^n)}}\right)\right) \ ,$$

for some constant D and $d = a^{1-\frac{1}{b}}$.

Proof. Let us fix constants $C > 1$, $c > 0$ such that $0 < c < |w(n)| \leq C^n$ for sufficiently large $n \in \mathbb{N}$. Then

$$\sum_{k=0}^{n} a^{(b^k)}w(k)\binom{n}{k}x^k(1-x)^{n-k} = \\ a^{(b^n)}x^n w(n)\left(1 + \sum_{k=0}^{n-1} a^{b^k - b^n}\frac{w(k)}{w(n)}\binom{n}{k}\left(\frac{1-x}{x}\right)^{n-k}\right) \ .$$

Let $y = \max\{1, \frac{1-x}{x}\}$. Then $\left(\frac{1-x}{x}\right)^{n-k} \leq y^n$ for each $k \in \{0, n-1\}$. Hence, for sufficiently large n, we have

$$\left|\sum_{k=0}^{n-1} a^{b^k - b^n}\frac{w(k)}{w(n)}\binom{n}{k}\left(\frac{1-x}{x}\right)^{n-k}\right| \leq \\ \sum_{k=0}^{n-1} a^{b^k-b^n}\frac{C^k}{c}2^n y^n \leq \frac{1}{c}\frac{n(2yC)^n}{a^{b^n - b^{n-1}}} = \\ \frac{1}{c}\frac{n(2yC)^n}{(a^{1-1/b})^{b^n}} = \mathrm{O}\left(\frac{D^n}{d^{(b^n)}}\right) \ ,$$

where $d = a^{1-1/b}$ and D is any real number bigger than $2yC$.

3.3 Bernstein Polynomials The Bernstein approximation B_n^f to a function $f : [0;1] \to \mathbb{R}$ (see, e.g., [10]) is the polynomial

$$B_n^f(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \ .$$

It is clear that $B_n^f(x)$ is the expected value of the random variable $f\left(\frac{X_n}{n}\right)$, where X_n is a random variable with the binomial distribution with parameters n an x. A direct application of Theorem 2.2 applied to Bernoulli trials and the function $g(x) = f(\frac{x}{n})$ shows the following result, proved by S. Bernstein in [11]: if $f^{(2s)}$ is bounded in $(0,1)$ and μ_a denotes ath central moment of a random variable with $\mathrm{Bin}(n,x)$ distribution, then

$$B_n^f(x) = f(x) + \sum_{a=2}^{2s-1}\frac{f^{(a)}(x)}{a!n^a}\mu_a + \mathrm{O}\left(\frac{1}{n^s}\right) \ .$$

This formula was used in [4] in order to give alternative, but longer than based on the Delta Method, proofs of results from papers [1], [2] and [3].

4 Probabilistic Poissonization

Let $f : \mathbb{R} \to \mathbb{R}$ and let X be a random variable with Poisson distribution with parameter λ. We put $\mathcal{P}(f; \lambda) = \mathrm{E}[f(X)]$, i.e., $\mathcal{P}(f; \lambda) = \mathrm{e}^{-\lambda}\sum_{k=0}^{\infty} f(k)\lambda^k/k!$ (see [12]).

The central moments of Poisson distribution with parameter λ satisfies the following recurrence $\mu_{k+1} =$

$\lambda\left(\frac{d\mu_k}{d\lambda}+k\mu_{k-1}\right)$. From this recurrence we deduce that μ_k is a polynomial in λ of order $\lfloor k/2\rfloor$. A direct application of Theorem 2.1 with $s=3$, $m=0$, $a=\lambda/2$, $b=M=\infty$ gives us the following formula:

$$\mathcal{P}(f;\lambda)=f(\lambda)+ \lambda\left(\frac{1}{2}f^{(2)}(\lambda)+\frac{1}{6}f^{(3)}(\lambda)+\frac{1}{24}f^{(4)}(\lambda)+\frac{1}{120}f^{(5)}(\lambda)\right)+ \lambda^2\left(\frac{1}{8}f^{(4)}(\lambda)+\frac{1}{12}f^{(5)}(\lambda)\right)+R_{6;\frac{\lambda}{2},\infty}\,,$$

where

$$|R_{6;\frac{\lambda}{2},\infty}|\le A\mathrm{M}^{(6)}_{(\frac{\lambda}{2},\infty)}(f)+B\sqrt{\Pr(X<\frac{\lambda}{2})}+ \mathrm{M}^{(0)}_{[0,\frac{\lambda}{2}]}(f)\Pr(X<\frac{\lambda}{2}).$$

Let us assume that $f(x)=\mathrm{O}(x^a)$ for some $a>0$. Then the last two terms of this estimation are exponentially small when λ grows to infinity (more precisely, $\Pr[X<\lambda/2]=\mathrm{O}\left(\sqrt{2/\mathrm{e}}^{\lambda}\right)$).

Let us apply this formula for entropy of the Poisson distribution. Hence, we consider the function $h(x)=-\log(\mathrm{e}^{-\lambda}\frac{\lambda^x}{\Gamma(x+1)})$. Clearly $h(x)=\mathrm{O}\left(x^2\right)$. Therefore, the last two terms from error term are exponentially small, hence they are of order $\mathrm{O}\left(\lambda^{-2}\right)$. We have $h^{(6)}(x)=\psi^{(5)}(x+1)$, where $\psi^{(5)}(x)$ denotes the polygamma function of order 5. Hence $\sup\{|h^{(6)}(x)|:x>\lambda/2\}=\mathrm{O}\left(\frac{1}{\lambda^5}\right)$. Therefore,

$$\mathrm{E}\left[h(X)\right]=\log\sqrt{2\pi\mathrm{e}\lambda}-\frac{1}{12\lambda}+\mathrm{O}\left(\frac{1}{\lambda^2}\right)+\mathrm{O}\left(\frac{1}{\lambda^5}\right)\mathrm{O}\left(\lambda^3\right)\,.$$

Hence we get the formula

$$H_\lambda=\log\sqrt{2\pi\mathrm{e}\lambda}-\frac{1}{12\lambda}+\mathrm{O}\left(\frac{1}{\lambda^2}\right)$$

for the entropy of Poisson distribution (see [13]). More precise approximation of H_λ can be derived in a similar way and the computation can be completely automated. In a similar way we can derive a formula from Corollary 3 from [3] for negative moments of Poisson random variables.

The next observation allows us to extend the class of functions for which we may use the Delta Method for the Poisson distribution.

LEMMA 4.1. *Let $a>0$ and $f(z)=w(z)\cdot a^z$. Then*

$$\mathcal{P}(f;\lambda)=e^{\lambda(a-1)}\mathcal{P}(w;\lambda a)\ .$$

Proof. Let $a>0$ and $\lambda>0$. Then

$$\mathcal{P}(f;\lambda)=e^{-\lambda}\sum_{k\ge 0}w(k)a^k\frac{\lambda^k}{k!}= e^{-\lambda}\sum_{k\ge 0}w(k)\frac{(\lambda a)^k}{k!}=e^{-\lambda}e^{\lambda a}\mathcal{P}(w;\lambda a)\ .$$

Let us observe that there is no analogous version of Lemma 3.2 for Poisson distribution. Namely, if $a>1$ and $\lambda>0$, then $\sum_{k\ge 0}a^{a^k}\frac{\lambda^k}{k!}=\infty$.

5 Negative Binomial Distribution

If X is a random variable with $\mathrm{NB}(n,p)$ distribution and $f:\mathbb{R}\to\mathbb{R}$ is any function, then we put

$$\mathcal{NB}(f;n,p)=\mathrm{E}\left[f(X)\right]=\sum_{k\ge n}f(k)\binom{k-1}{n-1}p^n(1-p)^{k-n}\,.$$

Let us recall that if $X\sim\mathrm{NB}(n,p)$, then there are independent random variables $Z_1,\ldots,Z_n$ with $\mathrm{Geo}(p)$ distributions such that $X=Z_1+\ldots+Z_n$.

A direct application of Theorem 2.2 with parameters $s=3$, $a=\frac{n}{2p}$ and $b=\infty$ gives us the following formula:

$$\begin{aligned}\mathcal{NB}(f;n,p)&=f\left(\frac{n}{p}\right)+\\ &\frac{n^2(1-p)^2}{24p^5}\left(3pf^{(4)}\left(\frac{n}{p}\right)+4f^{(5)}\left(\frac{n}{p}\right)-2pf^{(5)}\left(\frac{n}{p}\right)\right)+\\ &\frac{n(1-p)}{120p^5}\left(60p^3f''\left(\frac{n}{p}\right)+40p^2f^{(3)}\left(\frac{n}{p}\right)-20p^3f^{(3)}\left(\frac{n}{p}\right)+\right.\\ &30pf^{(4)}\left(\frac{n}{p}\right)-30p^2f^{(4)}\left(\frac{n}{p}\right)+5p^3f^{(4)}\left(\frac{n}{p}\right)+\\ &24f^{(5)}\left(\frac{n}{p}\right)-36pf^{(5)}\left(\frac{n}{p}\right)+\\ &\left.14p^2f^{(5)}\left(\frac{n}{p}\right)-p^3f^{(5)}\left(\frac{n}{p}\right)\right)+\\ &R_{6;\frac{n}{\sqrt{p}},\infty}.\end{aligned}$$

The Chernoff bound adjusted to the variables with negative binomial distribution gives us $\Pr\left(X\notin\left(\frac{n}{\sqrt{p}},\infty\right)\right)=\mathrm{O}\left((A_p)^n\right)$ for some $A_p<1$. Hence both terms $n^3B\sqrt{\Pr\left(X\notin\left(\frac{n}{\sqrt{p}},\infty\right)\right)}$ and $\mathrm{M}^{(0)}_{[0,\frac{n}{\sqrt{p}}]}(f)\Pr\left(X\notin\left(\frac{n}{\sqrt{p}},\infty\right)\right)$ are negligible for a function f of at most polynomial growth. Therefore

$$R_{6;\frac{n}{\sqrt{p}},\infty}=\mathrm{O}\left(n^3\mathrm{M}^{(6)}_{[0,\frac{n}{\sqrt{p}}]}(f)\right).$$

Let us apply this formula for entropy of negative binomial distribution. We have to consider the function $h(x)=-\log\left(\binom{x-1}{n-1}p^n(1-p)^{x-n}\right)$. Since $h^{(6)}(x)=\psi^{(5)}(x+1-n)-\psi^{(5)}(x)$, where $\psi^{(5)}(x)$ denotes the fifth order polygamma function, we have $sup\{|h^{(6)}(x)|:x>\frac{n}{\sqrt{p}}\}=\mathrm{O}\left(\frac{1}{n^5}\right)$. Therefore $R_{6;\frac{n}{\sqrt{p}},\infty}=\mathrm{O}\left(\frac{1}{n^2}\right)$ and the entropy of the negative binomial distribution is

$$H_{\mathrm{NB}(n,p)}=\log\sqrt{2\pi\mathrm{e}\frac{n(1-p)}{p^2}}-\frac{(2-p)^2}{12n(1-p)}+\mathrm{O}\left(\frac{1}{n^2}\right).$$

Using the symbolic calculation packages and calculating the ath central moment of the negative binomially distributed random variable from its central moments generating function or from the recurrence relation for the central moments ($\mu_a = q\frac{\partial \mu_{a-1}}{\partial q} + an\frac{q}{p^2}\mu_{a-2}$, where $q = 1-p$, see [14]) one can obtain an approximation of $H_{\mathrm{NB}(n,p)}$ with an arbitrary required precision.

The next observation allows us to extend the class of functions for which we may extend the Delta Method for the negative binomial distributions.

LEMMA 5.1. *Let* $0 < a < \frac{1}{1-p}$ *and* $f(z) = w(z)\cdot a^z$. *Then*

$$\mathcal{NB}(f;n,p) = \left(\frac{ap}{1-a(1-p)}\right)^n \mathcal{NB}(w;n,1-a(1-p)) \cdot$$

Proof. Let $q = 1-p$. Then

$$\left(\frac{ap}{1-a(1-p)}\right)^n \mathcal{NB}(w;n,1-a(1-p)) =$$
$$\left(\frac{ap}{1-aq}\right)^n \sum_{k\geq n} w(k)\binom{k-1}{n-1}(1-aq)^n (aq)^{k-n} =$$
$$\sum_{k\geq n} w(k)a^k \binom{k-1}{n-1} p^n (1-p)^{k-n} .$$

Remark. If $a \geq \frac{1}{1-p}$ and $f(z) = a^z$, then the series $\mathcal{NB}(f;n,p)$=$\sum_{k\geq n} a^k \binom{k-1}{n-1} p^n (1-p)^{k-n}$ is divergent.

6 Hypergeometric Distribution

A random variable X has a hypergeometric $\mathrm{HGeo}(n,n_1,n_2)$ distribution with parameters n, n_1 and n_2 if $\Pr(X=k) = \binom{n_2}{k}\binom{n-n_2}{n_1-k}\Big/\binom{n}{n_1}$. The random variable X counts the number of successes in n_1 draws without replacement from a finite population of size n containing n_2 success states.

We fix two parameters $p,q \in (0,1)$ and consider a random variable X with distribution $\mathrm{HGeo}(n,pn,qn)$. Note that $\mathrm{E}[X] = npq$. Without loss of generality we assume that $p \leq q$ and $p+q \leq 1$ (one can show correctness of the result without the additional assumptions, by separately considering the remaining cases or by taking into account symmetries of hypergeometric distribution). Our goal is to derive an asymptotic formula for the entropy of random variable X, when n tends to infinity. We apply Theorem 2.1 to the function $h(k) = -\log\left(\binom{qn}{k}\binom{n-qn}{pn-k}\Big/\binom{n}{pn}\right)$. We have

$$h^{(a)}(x) = \psi^{(a-1)}(1+x) + \psi^{(a-1)}(1+(1-p-q)n+x)$$
$$+(-1)^a\psi^{(a-1)}(1+pn-x) + (-1)^a\psi^{(a-1)}(1+qn-x).$$

We set $m=0$, $M = pn$, $a = \frac{1}{2}pqn$ and $b = \frac{1+q}{2}pn$. As probability $\Pr(X \notin (a,b))$ is exponentially small (see, e.g., [15]), we focus on the first part of remainder. We have $\mu_{2s}(X) = \mathrm{O}(n^s)$ and $\mathrm{M}^{(2s)}_{(a,b)}(h) = \mathrm{O}\left(n^{-2s+1}\right)$. Thus, in order to obtain result up to $\mathrm{O}\left(\frac{1}{n^2}\right)$ term, we take $s=3$ and after some calculations we get

$$H_{\mathrm{HGeo}(n,pn,qn)} = \log\sqrt{2\pi n e p^* q^*} +$$
$$\frac{1}{12n}\Big(-10 + \frac{4}{p^*} + \frac{4}{q^*} - \frac{1}{p^* q^*}\Big) + \mathrm{O}\left(\frac{1}{n^2}\right) ,$$

where $p^* = p(1-p)$ and $q^* = q(1-q)$.

7 Conclusions

Properly formulated Delta Method like theorems can accelerate and simplify determination of approximations of probability sums for the distributions that are well concentrated near its mean value for large class of functions. Methods presented in this paper can be automated to a large extent.

References

[1] P. Jacquet, W. Szpankowski, Entropy computations via analytic depoissonization, IEEE Transactions on Information Theory 45 (4) (1999) 1072–1081.

[2] P. Flajolet, Singularity analysis and asymptotics of bernoulli sums, Theor. Comput. Sci. 215 (1-2) (1999) 371–381.

[3] M. Žnidarič, Asymptotic expansion for inverse moments of binomial and poisson distributions, The Open Statistics & Probability Journal 1 (2009) 7–10.

[4] J. Cichoń, Z. Gołębiewski, On bernoulli sums and bernstein polynomials, DMTCS Proceedings (2012) 179–190.

[5] H. Cramér, Mathematical methods of statistics, Princeton University Press, 1946.

[6] G. W. Oehlert, A Note on the Delta Method, The American Statistician 46 (1) (1992) 27–29. doi:10.2307/2684406.

[7] J. Cichon, J. Lemiesz, W. Szpankowski, M. Zawada, Two-phase cardinality estimation protocols for sensor networks with provable precision, in: WCNC, IEEE, 2012, pp. 2009–2013.

[8] P. Bickel, K. Doksum, Mathematical statistics: basic ideas and selected topics, no. t. 1 in Mathematical Statistics: Basic Ideas and Selected Topics, Prentice Hall, 2001.

[9] J. Marcinkiewicz, A. Zygmund, Sur les foncions independantes, Fund. Math. 28 (1937) 60–90.

[10] G. Lorentz, Bernstein polynomials, Chelsea Publishing Series, Chelsea Pub. Co., 1986.

[11] S. Bernstein, Complement a l'article de E. Voronovskaya, C. R. Acad. URSS (1932) 86–92.

[12] P. Jacquet, W. Szpankowski, Analytical depoissonization and its applications, Theor. Comput. Sci. 201 (1-2) (1998) 1–62.

[13] J. A. Adell, A. Lekuona, Y. Yu, Sharp bounds on the entropy of the poisson law and related quantities, CoRR abs/1001.2897.

[14] V. N. L. Johnson, A. W. Kemp, S. Kotz, Univariate Discrete Distributions, New York: John Wiley, 1992.

[15] R. J. Serfling, Probability inequalities for the sum in sampling without replacement, Ann. Statist. 2 (1) (1974) 39–48.